PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : G06F 11/20, 11/16 — A1

(11) International Publication Number: WO 93/23805

(43) International Publication Date: 25 November 1993 (25.11.93)

(21) International Application Number: PCT/US93/04444

(22) International Filing Date: 11 May 1993 (11.05.93)

(30) Priority data:
881,434 11 May 1992 (11.05.92) US

(71) Applicant: TANDEM COMPUTERS INCORPORATED [US/US]; 10435 N. Tantau Avenue, Loc. 200-16, Cupertino, CA 95014-0709 (US).

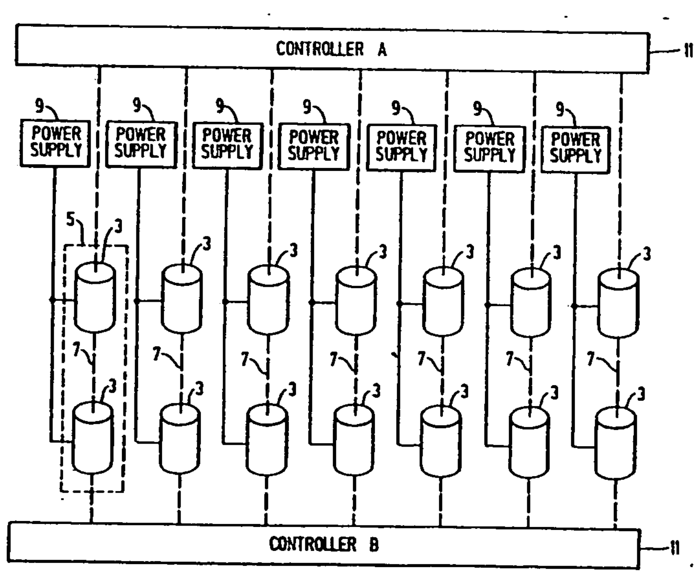
(72) Inventor: LUI, Albert, S. ; 3164 Heritage Valley Drive, San Jose, CA 95148 (US).

(74) Agents: LAND, John et al.; Spensley-Horn Jubas & Lubitz, 1880 Century Park East, Fifth Floor, Los Angeles, CA 90067 (US).

(81) Designated States: JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
With amended claims.

(54) Title: FAULT TOLERANT POWER SUPPLY FOR AN ARRAY OF STORAGE DEVICES



(57) Abstract

A fault tolerant power supply system for providing reliable power to a redundant array of channel storage units. The system includes one power supply module (9) for each channel (5) of the array of data storage units (3). A power supply failure will not impact the ability of the data storage system to recover data due to the ability of the data storage system to reconstruct data in an unavailable channel (5) from the data storage units (3) of each other channel (5). The use of independent power supplies (9) provides a power supply system which has a power capability equal to the sum of the power requirements of the data storage units (3), and voltage outputs just sufficient to meet the voltage requirements of the data storage units (3).

FAULT TOLERANT POWER SUPPLY FOR AN ARRAY OF STORAGE DEVICES

BACKGROUND OF THE INVENTION

1. *Field of the Invention*

This invention relates to fault tolerant power supplies and more particularly to fault tolerant power supplies for supplying electrical power to a redundant array of data storage units.

2. *Description of Related Art*

A typical data processing system generally includes one or more storage units connected to a Central Processor Unit (CPU) either directly or through a control unit. The storage units store data and programs which the CPU uses in performing particular data processing tasks.

Various types of storage units are used in current data processing systems. A typical system may include one or more large capacity tape units and/or disk drives (magnetic, optical, or semiconductor) connected to the system through respective control units for storing data. In such systems, a problem exists if one of the large capacity storage units fails such that information contained in that unit is no longer available to the system. Often, such a failure will shut down the entire computer system. Such a failure may arise from a defect in the storage unit, or from a fault in the power supply for the storage unit. Therefore, among other things, it is critical for the power supply for such data storage units to be fault tolerant. It is also critical to be able to recover the data stored in a data storage unit if the storage unit fails for any reason, including loss of power.

A RAID 3 architecture is based on the concept that each disk drive storage unit has internal means for detecting a fault or data error. Therefore, it is not necessary to store extra information to detect the location of an error; a simpler form of parity-based error correction can thus be used. In this approach, the contents of all storage units subject to failure are "Exclusive OR'd" (XOR'd) to generate parity information. The resulting parity information is stored in a single redundant storage unit. If a storage unit fails, the data on that unit can be reconstructed onto a replacement storage unit by XOR'ing the data from the remaining storage units with the parity information.

A RAID 4 architecture uses the same parity error correction concept of the RAID 3 architecture, but improves on the performance of a RAID 3 system with respect to random reading of small files by "uncoupling" the operation of the individual disk drive actuators, and reading and writing a larger minimum amount of data (typically, a disk sector) to each disk (this is also known as block striping). A further aspect of the RAID 4 architecture is that a single storage unit is designated as the parity unit.

A RAID 5 architecture uses the same parity error correction concept of the RAID 4 architecture and independent actuators, but improves on the writing performance of a RAID 4 system by distributing the data and parity information across all of the available disk drives.

All of the RAID architectures use multiple data storage units. As discussed above, in addition to ensuring that the failure of one of the data storage units will not cause a loss of stored data, supplying the power requirements for a plurality of storage units is critical to proper operation of the data processing system. In many data processing systems, the supply of reliable power to the data storage units is ensured by use of redundant power supplies, each of which is capable of supplying the electrical power requirements of all the data storage units.

Furthermore, it is generally essential for the voltage supplied to the data storage units to remain constant. Since the voltage drop of the blocking diodes 2 varies as a function of the current through each diode, and the current requirements of each data storage unit 3 vary as a function of time, the voltage supplied to each storage unit 3 will vary as a function of time unless regulated.

One voltage regulator 4 dedicated to each storage unit 3 is provided to ensure that the proper voltage is maintained at each storage unit 3 (See FIGURE 1A). Because voltage regulation circuits generally require a voltage higher than the stable output desired, each redundant power supply must provide a voltage level higher than the voltage level that would be required if local voltage regulation were not needed. Therefore, the redundant power supplies must be larger than would be necessary in the absence of such local voltage regulators within each storage unit. Additionally, each regulation circuit will add to the overall cost of the system.

Alternatively, a sensing circuit that dynamically determines the amount of voltage lost due to the diodes may be used to provide feedback to an active voltage regulator circuit that adjusts the power supply output voltage to compensate for these losses. Such circuits are known. The obvious disadvantage to this approach is the need for additional circuitry.

Therefore, it is desirable to create a fault tolerant power supply system that has a power capability that need not be greater than the electrical power required by the sum of all the storage units during normal operation, and which has a voltage output just sufficient to meet the voltage requirements of the data storage units.

sufficient to meet the voltage requirements of the data storage units. No blocking diode or additional voltage regulation circuitry is required at the data storage device to achieve fault tolerance.

The details of the preferred embodiment of the present invention are set forth in the accompanying drawings and the description below. Once the details of the invention are known, numerous additional innovations and changes will become obvious to one skilled in the art.

DETAILED DESCRIPTION OF THE INVENTION

Throughout this description, the preferred embodiment and examples shown should be considered as exemplars, rather than as limitations on the present invention.

Referring to FIGURE 2, the inventive power supply system is shown used in a data storage system 1 including an array of data storage units. Channels of data storage units 3 are shown coupled by broken vertical lines representing I/O buses 7 (such as Small Computer System Interface Buses or SCSI Buses). A channel 5 of data storage units 3 is defined as all data storage units 3 in the array coupled by one I/O bus 7. Each channel 5 is connected to at least one data controller 11 that controls Read and Write operations which transfer data to and from the data storage units 3. The array of data storage units 3 shown can function as a RAID 1, RAID 2, RAID 3, RAID 4, or RAID 5 system. In any case, the data stored is recoverable upon a failure of any number of data storage units 3 of any one channel 5 using known RAID recovery techniques.

A single power supply 9 is coupled to each channel 5 of data storage units 3 and supplies the power requirements of that channel 5. Each power supply 9 has a maximum power output rating at least equal to the maximum power requirement of the sum of the data storage units 3 in the corresponding channel 5.

The data stored in the data storage units 3 of a single channel 5 can be recovered upon failure of the entire channel 5, using known RAID recovery techniques. Thus, failure of any one power supply 9 will not cause data to become unavailable to a requesting device, such as a controller 11, even though the entire corresponding channel 5 of data storage units 3 will become inoperable upon such a failure.

embodiment is that there is no need for blocking diodes. Therefore, the voltage output of the selected power supply 9' need not be higher than the voltage input requirement of the attached controller 11 and data storage units 3.

Accordingly, since a failure of any one channel of storage units in a RAID system (whether the data storage units are configured in a RAID 1, RAID 2, RAID 3, RAID 4, or RAID 5 configuration) can be tolerated without any loss of data, the modular architecture of the inventive power supply system makes it unnecessary for the power supply system capability to exceed the power requirements of the components to which power is supplied. Also the modular architecture eliminates the need for blocking diodes. Therefore, it is unnecessary for the power supply output voltage to exceed the voltage requirements of the components to which power is supplied.

A number of embodiments of the present invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, it is to be understood that the invention is not to be limited by the specific illustrated embodiment, but only by the scope of the appended claims.

6. The fault-tolerant power supply system of claim 1, wherein the redundant array of failure independent data storage units are arranged as a RAID 1 system, such that for each data storage unit present in the redundant array having a set of stored data, there is one corresponding data storage unit having stored data, the stored data of each data storage unit being identical to the stored data of the corresponding data storage unit.

7. The fault-tolerant power supply system of claim 1, wherein the redundant array of failure independent data storage units are arranged as a fault tolerant RAID redundancy group including any number of data storage units, each from a distinct channel, each of which stores data which can be used to reconstruct data stored in any one data storage unit within the redundancy group.

4. The fault-tolerant power supply system of claims 2 or 3, wherein the independent power supplies have a combined power output capability that is substantially equal to the combined maximum power requirements of all the coupled data storage units.

5. The fault-tolerant power supply system of claims 2 or 3, wherein the redundant array of failure independent data storage units are arranged as a RAID 1 system, such that for each data storage unit present in the redundant array having a set of stored data, there is one corresponding data storage unit having stored data, the stored data of each data storage unit being identical to the stored data of the corresponding data storage unit.

13. The fault-tolerant power supply system of claims 4, wherein the independent power supplies have a combined power output capability that is substantially equal to the combined maximum power requirements of all the coupled data storage units.

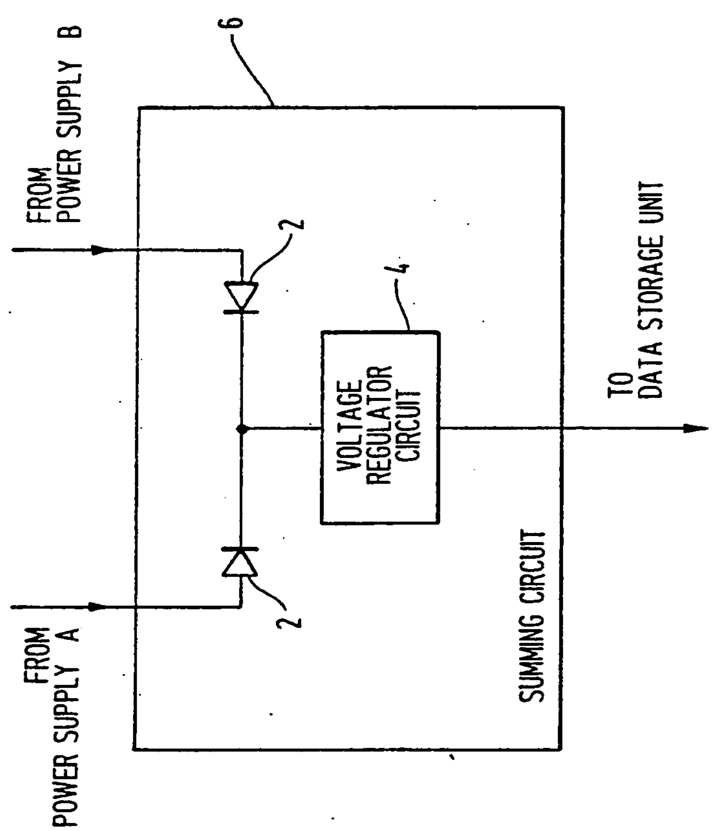


FIG. 1A
PRIOR ART

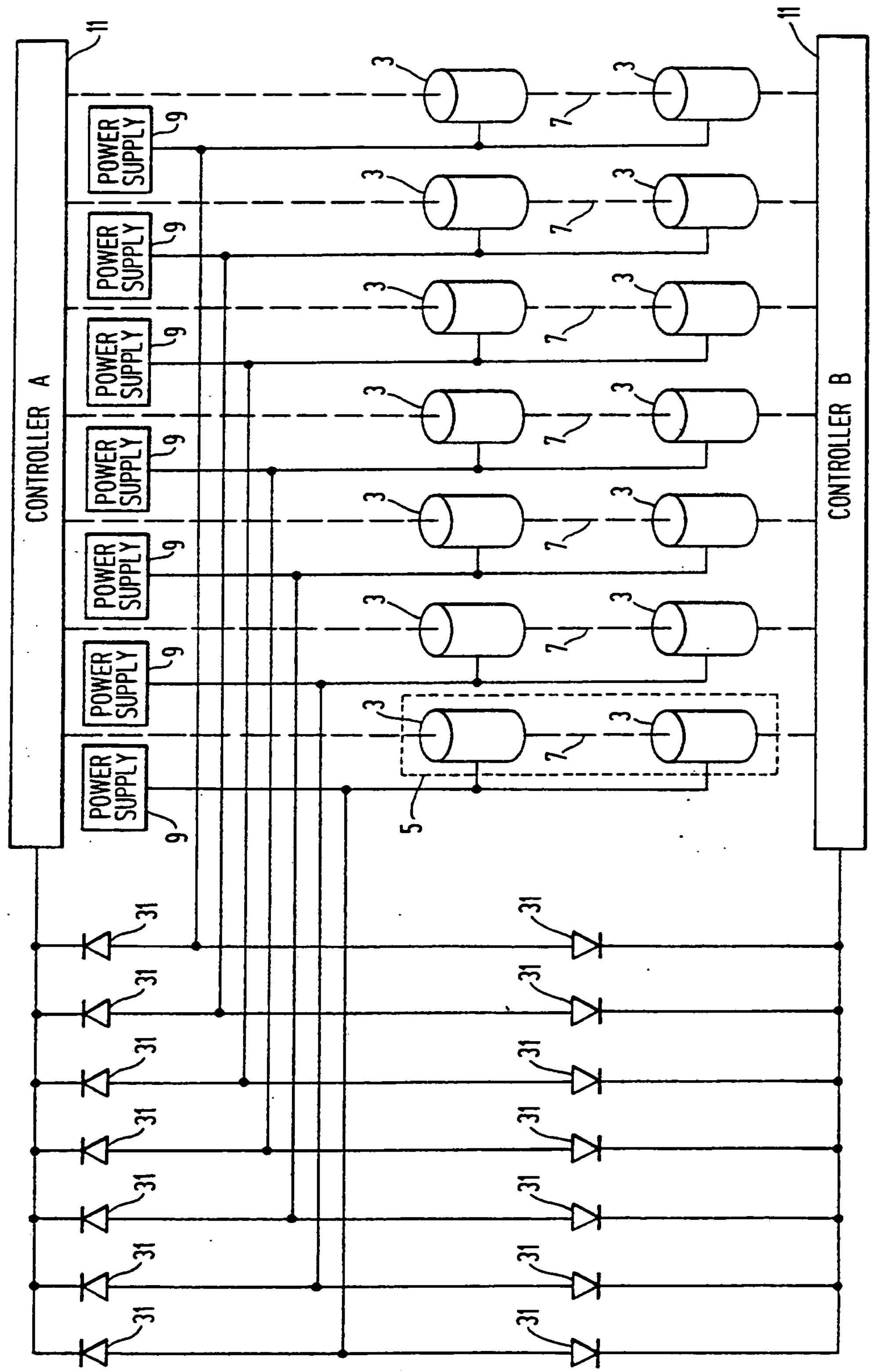


FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/04444

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :G06F 11/20, 11/16
US CL :371/8.1, 10.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 371/8.1, 10.1, 66

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS: subclasses (364/(246.4 or 268.5 or 966.5) and 364/(273.3 or 273.4 or 948.5 or 948.7)); array (2a) disk? and classes (360 or 364 or 371 or 395) and power.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Schulze, M., "Considerations in the Design of a RAID Prototype", Master's Report Number UCB/CSD 88/448, published 1988 by Computer Science Division (EECS), University of California (Berkeley, California, U.S.A.), see figure 2, lines 20-22 on page 14, and lines 5-10 on page 18.	1, 4-7
X	US, A, 4,754,397 (Varaiya et al) 28 June 1988, see column 5 at lines 37-42, column 6 at lines 60-65, column 7 at lines 15-19, figures 4 and 7, abstract.	1, 4-7

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 04 August 1993	Date of mailing of the international search report 24 AUG 1993
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. NOT APPLICABLE	Authorized officer STEPHEN M. BAKER Telephone No. (703) 305-9600